

---

# A Study of Mate Selection Schemes in Genetic Algorithms—Part I

**Chien-Feng Huang**

Modeling, Algorithms, and Informatics Group (CCS-3)  
Computer and Computational Sciences  
Los Alamos National Laboratory, MS B256  
Los Alamos, NM 87545, USA  
cfhuang@lanl.gov

---

## Abstract

In the Genetic Algorithm (GA) literature, many models focus on problems where each individual's fitness is independent of others (or implicitly defined by others). In this paper, a framework is introduced for studying mate selection in the context of GA. The objective is to model interdependent fitnesses of population individuals by allowing them to search for mates. The resulting GA thus forms a more complex system in which each individual's fitness depends on both the environment and other population members. The methods for investigation consist of the Schema Theorem, a Markov chain analysis, along with several empirical results. I will show that mate selection plays a crucial role in GA's search power. In particular, individuals with more distinct characteristics collectively facilitate the search for a single, better solution. The results presented on the effects of mate selection are a first step toward a deeper understanding of how GAs work, and thus how to design more robust GAs.

## Keywords

Genetic algorithms, mate selection, Schema Theorem, hitchhiking, founder effect, Markov chain analysis, similar mating, dissimilar mating.

## 1 Introduction

### 1.1 Motivation

A Simple GA (Mitchell, 1996) traditionally generates a random, initial population of candidate solutions (chromosomes). Selection for reproduction, crossover and mutation operate on the population over a certain number of generations until a stop criterion is reached. The probability of individuals being selected for reproduction is based on their fitness values: better solutions have larger probability to be chosen to cross with other solutions and generate offspring that share the genetic material from both parents. Mutations may occur with very low probability. If there is no specific restriction on how mating partners should be chosen, an individual mates with any other regardless of its parenthood or likeness. This type of mating scheme is referred to as *random mating* (Roughgarden, 1979; Russel, 1998).

Random mating is the simplest form of mating process used by GAs. The selection scheme acts on the population to pick two parents for producing offspring, each parent being *passively* assigned its mating partner. Such a simple implementation in fact overlooks a potential advantage in the GA's mating process—the design of robust GAs is contingent not only upon how they exchange information (the task of crossover),

but also upon their selection of proper mates (the task of mate selection). It may be quite beneficial to seek a mate by active search, rather than by happenstance; once a good mate is found, the information exchange and combination may be facilitated to improve performance substantially. Consider the following example:

Suppose the population is composed of bit-strings of length 8 and the relevant building blocks are 1111\*\*\*\* and \*\*\*\*1111 (\* can be either 1 or 0), each of them contributing fitness of 4 to the strings in the population. Then, for example, a string  $X$ , 11110000, is of fitness 4, and the optimal string is 11111111, whose fitness is 8. If we are given string  $X$  and two candidate mating partners,  $Y_1$  (11110000) and  $Y_2$  (00001111), under the mechanism of a simple GA (Mitchell, 1996),  $Y_1$  and  $Y_2$  have the same probability to be chosen for mating since these two strings are of the same fitness. However, if we are concerned with finding the optimal string, clearly, string  $Y_2$  is better than  $Y_1$  because the mating between  $Y_2$  and  $X$  is likely to generate the optimum, but it is not the case if  $X$  mates with  $Y_1$ . This implies that the simple selection scheme is not able to distinguish individuals of the same fitness, yet of quite different string structures.

The example above shows a potential deficiency of employing a simple selection scheme that does not permit individuals to *actively* determine their mates. Such an implementation of the selection process confers the only selection pressure in simple GAs, where population members' fitnesses are independent of others in the sense that their fitnesses are explicitly determined by the environment, rather than by other individuals (or implicitly determined by others). In this paper, I introduce another source of selection pressure by allowing individuals themselves to actively choose mates. The fitnesses of candidate mates are then somewhat "re-defined" according to the degree of their satisfying mating preference of a given individual. As a result, the fitnesses of population members can be explicitly interdependent and coevolve with each other.

Before delving fully into this paper, it is important to discuss biological background of non-random mating, and provide a literature review of prior GA research work on this subject.

## 1.2 Biological Background

One of the reasons population genetics is a successful theory is that it is built upon a null model, the Hardy-Weinberg equilibrium principle (Freeman and Herron, 1998, p. 121). This null model predicts, under certain assumptions (i.e., population experiencing no selection, no mutation, no migration, no genetic drift and random mating), that across generations allele and genotype frequencies will not change. The Population will not evolve. Violations of these simple assumptions of the null model can result in change of allele frequencies. Population genetics thus identifies the mechanism of population evolution.

The assumption of the Hardy-Weinberg Law that members of opposite sexes in populations mate randomly ensures that the population equilibrium is not disturbed. This case is referred to as *panmixia* in which each individual member of a population has an equal chance of mating with every other individual of the opposite sex. The probability of mating is not determined by genotype; each individual mates without preference. However, any degree of preferential matings between individuals that results in particular genotypes mating more frequently than would be expected at random constitutes non-random mating. When non-random mating occurs, genotypes will not occur in frequencies predicted by the Hardy-Weinberg equilibrium principle.

Sexual selection is typical of non-random mating, where members of one sex show a consistent preference for a particular *phenotype* of the opposite sex (Price, 1996; Free-

man and Herron, 1998). It normally results in strongly skewed mating systems, because panmixis is lost and often a small number of individuals are responsible for much of the reproduction in a population.

Another representative non-random mating mechanism is assortative mating (Price, 1996; Freeman and Herron, 1998). Assortative mating means mating within subgroups of a population characterized by genetic similarities between mates. For example, organisms with similar genetically controlled heat requirements tend to become sexually active at the same time and mate with each other more frequently than with individuals with different heat requirements.

Assortative mating can be generally classified as four types: *positive assortative mating*, *negative assortative mating*, *inbreeding*, and *outbreeding*. Positive assortative mating indicates that individuals preferentially mate with their own phenotype so as to decrease heterozygotic frequency (increases homozygotic frequency). This is common in humans; for example, tall men and tall women tend to marry. Negative assortative mating, on the contrary, means that individuals avoid their own phenotype and the heterozygotic frequency is increased. For example, if tall persons select short persons to marry this would represent negative assortative mating.

The third type is *inbreeding* or, more commonly, *incest* (Roughgarden, 1979; Russell, 1998): close relatives have an increased probability of mating with each other, and unrelated individuals have a decreased probability of mating. *Outbreeding* refers to less mating between relatives than occurs in random mating. Inbreeding increases the proportion of homozygous individuals in a population at the expense of heterozygotes, thereby resulting in decrease of the genetic diversity in the population. Outbreeding, on the other hand, increases that same diversity (Roughgarden, 1979). In case of inbreeding, genetic disorders oftentimes increase in frequency with increase in homozygous recessives. *Inbreeding depression* is the reduction in mean fitness of a population from inbreeding resulting in homozygotes with deleterious or lethal recessive alleles. This is a significant problem in zoos and captive breeding programs, as well as for endangered species such as the California condor.

### 1.3 Non-Random Mating in Prior GA Research

Biological background of non-random mating (restricted mating) provides significant insights for deeper understanding of the role of sex in evolution. In prior GA research, several non-random mating models are proposed, mainly aiming at (1) promoting population diversity, (2) preventing the formation of lethal solutions, and (3) inducing speciation.

Promoting population diversity for improving GA's performance of locating a single desired solution (e.g., the global optimum or the best-so-far individual in the population) is a common practice. Booker (1987) proposed crossover among reduced surrogates. If the population individuals differ in more than one bit, crossover is guaranteed to generate offspring that are different from their parents. His idea is to restrict mating between individuals that are too similar so as to prevent fixation within classes. In (Eshelman and Schaffer, 1991) *incest prevention* is another non-random mating scheme, where mating is allowed only between dissimilar individuals. If strings' Hamming distances exceed a certain threshold, they are permitted to mate. (Craighurst and Martin, 1995) also proposed a method for incest prevention in which recombination between individuals with a certain degree of shared parenthood is not permitted. The authors defined an incest prevention degree, which designates how far back in the family tree the GA must look in order to prevent recombination between two related individuals.

Fernandes, Tavares, Munteanu, and Rosa (2001) proposed the *negative assortative mating* GA that chooses one parent (the first parent) and a set  $A$  of individuals by regular selection schemes (e.g., fitness proportionate selection). Then the second parent will be the one, belonging to set  $A$ , that has the maximum Hamming distance to the first parent. If several second parents are of the same Hamming distance, the one with higher fitness is selected. The GAs with these “dissimilarity-based” mate selection schemes demonstrated improved performance when the goal is to locate a single desired solution.

Lethal solutions are the low-fitness offspring that may be generated from crosses between multiple conflicting solutions. For example, consider a function with two equal peaks:  $f(x) = (x - (\frac{1}{2}))^2$ , where 0 and 1 are the two optima. In many encodings, 00...0 and 11...1 may represent these two optima. The crossing from solutions at these two optima is rather likely to create useless hybrids that degrade online performance of GA. Resolving this problem usually involves a certain degree of restriction on mating, because the formation of lethal solutions can be avoided. Deb and Goldberg (1989) restricted mating by allowing an individual to search for a mate within a distance of  $\sigma_{mating}$ . If such an individual can be found, mating is performed. Otherwise, a random individual is chosen. They showed that restricting mating to similar individuals produces more consistent results across multiple runs, and improves average population fitness.

The third objective related to non-random mating is speciation—formation of reproductively isolated groups of organisms. In GA research, speciation mechanisms were conceived of mainly based on two classes of speciation, namely, allopatric speciation (Mayr, 1942) and sympatric speciation (Dobzhansky, 1937). Allopatric speciation imposes direct or indirect mate restriction on mating through geographic barriers. Canonical examples of direct restricting mating via geographic barriers are fine-grained parallel GAs (PGAs), where each individual is allocated at each deme and crossovers occur only between individuals that are near one another geographically. This is a form of local mating.

As opposed to the local mating algorithms, coarse-grained PGAs indirectly restrict mating in which the population is explicitly divided into smaller subpopulations. Each subpopulation is isolated from the others in the sense that it evolves independently with occasional migrations of individuals from one subpopulation to another. The resulting mating can only take place within geographically separate groups, except the migrating individuals. (See (Cantú-Paz, 1997) for a nice review regarding the existing PGA models.)

Sympatric speciation, on the other hand, restricts mating by explicit rules, rather than by geographic barriers. For example, Booker (1982, 1985) restricted matings to occur between functionally similar individuals. The individuals were rules in a Holland-style classifier system. Only classifiers that match (or partially match) the same message are allowed to cross. Along with a sharing scheme in selection, this approach was shown to work well for forming sub-populations in the context of classifier systems.

Booker (1982) and Goldberg (1989) explored various approaches in which a mating tag is attached to each individual. This tag must match another individual in some number of loci before a cross is permitted. Many variations exist, including one-way matching, two-way matching, and partial matching. More advanced methods add a template to each individual, and matches must occur between tags and templates instead of between tags and individuals. Tags and templates evolve, along with the rest of a string. Deb applied evolving species tags and templates to restrict mating in multimodal function optimization (Deb and Goldberg, 1989). Two individuals cross if their

tags and templates match both ways; if the matching is only one way, they cross with probability 0.5.

Spears (1994) employed a  $k$ -bit tag attached to each population individual, which effectively divides the population into  $2^k$  subpopulations. He then disallowed mating between individuals with different tags. Tags are never modified by mutation and crossover, yet selection is allowed to operate across subpopulations. The results he obtained showed that on two sinusoidal test functions, each having five peaks of equal height, the algorithm maintains subpopulations on multiple peaks. However, if peaks are of different heights, all subpopulations eventually converge to the highest peak.

Todd and Miller (1991) used mating-preference tag to induce sexual selection of individuals and demonstrated that sympatric speciation can be formed in their GA's framework. The tag decodes to a real number on a given interval that indicates an individual's preferred mating distance. The partial probability that a given individual mates with another individual is determined by a triangular function. If two individuals' partial probabilities are calculated, the probability that they actually mate is the product of their individual, partial probabilities.

With a specific selection-for-reproduction scheme, Todd and Miller showed that, on a flat fitness function, over the course of the run, the resulting population is dynamically divided into several clusters of phenotypically similar individuals: some subpopulations merge and some split into sub-clusters. However, on non-flat fitness functions, selection pressure and noise become dominant, and the GA does not generally achieve speciation.

Mahfoud (1995) proposed a scheme that restricts mating to take place only within a species; if two individuals are from different species, the crossing is not permitted, and the individuals proceed to the mutation stage. His implementation used a phenotypic distance threshold of 0.1 to decide whether two individuals belong to the same species. It turned out that runs on a sinusoidal testbed that has five peaks, spaced at intervals of 0.2, without mating restriction, the GA fully converges by generation 40 to a single global optimum from the five possibilities. (All 100 individuals become identical.) With mating restriction, due to the reduced number of crosses, the GA fully converges by generation 30 to a single global optimum. Mating restriction, in this case, accelerates convergence to a single peak, rather than distributing population on several peaks—speciation does not occur through this intraspecies mating restriction.

In conclusion, the prior existing work reveals that dissimilarity-based mating restriction seemingly facilitates the search for a single, satisfactory solution. Crosses between species appear to be desirable for potential jumps to higher peaks in the search space. However, to avoid producing lethal offspring, it would require that matings occur only between individuals of similar characteristics.

Speciation is a much more complicated issue. If selection pressure dominates the evolution process, similarity-based mating restriction alone generally does not guarantee effective speciation. The result Mahfoud obtained based on phenotypic similarity is a clear example. Several aforementioned similarity-based mating restrictions, including Spears' tag-based restriction and Todd and Miller's sexual selection (on non-flat fitness functions), are not successful in formation of species, either. To facilitate speciation would require additional strategies to control selection pressure. For instance, Deb and Goldberg (1989) employed the fitness sharing scheme to regulate fitnesses of population members based on the information available for niches in question. The results showed that their approach indeed promotes speciation and prevents convergence on a single peak. When similarity-based mating restriction was applied, the resulting spe-

ciation becomes more consistent and stable.

We may put together these observations and propose two claims as follows:

**Claim A:** Dissimilarity-based mating selection accelerates exploration of the search space.

**Claim B:** Similarity-based mating selection enhances the capability of speciation mechanism for formation and maintenance of subpopulations.

This means that there exists a delicate balance between mating like and unlike individuals if one aims at not only *rapid* formation, but also *stable* maintenance of a single solution or multiple solutions. Both interspecies and intraspecies crosses may be beneficial, depending on the problems at hand, the algorithms used, and the objectives.

Given all the results obtained so far in the context of GA, a systematic investigation on similarity and dissimilarity-based mate selection is still lacking. Todd and Miller (1997) pointed out that the primary reason was because traditional genetic algorithms research assumes that sexual recombination per se offers the major advantage of sexual reproduction (Goldberg, 1989). As such, one crucial issue concerning information exchange within a population of traditional GAs has been overlooked to some degree: the choice of mating partners can be critical. The goal in this paper is to propose a framework that allows us to model and investigate interdependent fitnesses of population members through their endogenous interactions. I will concentrate on claim A (see (Huang, 2002a) for some preliminary work on claim B) and show that mate selection plays a crucial role in GA's performance advancement.

## 2 The Schema Analysis and Empirical Results

Traditional GA models focus on problems in which each population member's fitness is independent of other population members—their fitnesses are evaluated by a static, fixed objective function. In nature the fitness of an individual depends in various ways on dynamic features. Organisms may interact and coevolve with other organisms in an environment. As a result, the fitnesses of organisms are evaluated not only by the environment but also by the individuals with whom they interact.

Several authors have studied GA models in which the fitness evaluation of individuals depends on other individuals in the same population or in other populations. Some of the results reported show that these models yield better performance than traditional optimization approaches (Hillis, 1992; Potter et al., 1995; Rosin and Belew, 1997).

In this paper, I introduce simple models that implement mate selection in the context of GA, based on assortative mating in biology. The idea is that, for instance, an individual may first choose a set of candidate mating partners. Then the probability of these candidates being selected as the actual mate may depend on the degree of their satisfying the first individual's mating preference. This method can be regarded as redefining the fitnesses of these candidates, and the probability of being selected is an increasing function of the newly defined fitnesses. As a result, the realized fitness of an individual depends on both the environment and how many others would consider it a good mate.

In the following subsections, the traditional fitness proportionate selection scheme is employed for two special cases of mating choices to facilitate a theoretical analysis. Although they represent two opposite extremes, it turns out that such an analysis can be applied to the investigation for some well known phenomena in biology and in

the context of GA, such as hitchhiking and the founder effect. These two phenomena have been identified as important factors that impede the GA's search power. The experimental results will show that a dissimilar mating choice effectively suppresses these two phenomena.

Next, I extend the range of mate selection to include two intermediate cases that combine similarity test and fitness-proportionate selection to further discriminate candidate mates. Therein it will be more clearly seen that the individuals' probabilities of being selected for mating are complicated by the environments and mate preferences. I then study the GA's performance in terms of "rate of improvement" and "creation of lethal hybrids." The results show that dissimilar mating schemes further the GA's exploration of the search space and yield a better best-so-far performance, yet at the expense of generating more lethal hybrids that degrade the population's fitness.

## 2.1 Mate Selection Schemes

Many approaches to implement the process of mate selection are possible. For example, the first individual may choose a set of members in the population, and then select the one who most matches its preference. If we adopt the Hamming distance as the similarity metric, then, in case of similar mating, the first individual may select as the actual mate the one whose Hamming distance to him is the smallest. In case of dissimilar mating, the one whose Hamming distance is the largest will be selected. Such mate selection schemes are deterministic in the sense that the individual who is finally being selected is the most similar (or dissimilar) to the first individual. It is quite natural to implement a probabilistic mate selection strategy. An example is to designate the probabilities of candidates being selected as the actual mate to be proportional to the magnitude of their Hamming distances.

In this paper, we replace "positive assortative mating" used in biology with "similar mating." Likewise, "negative assortative mating" is replaced with "dissimilar mating." The similarity measure between two individuals (bit strings) used here is Hamming distance—the number of locations at which corresponding bits differ.

Several mate selection schemes are proposed in this paper. We first examine two extreme cases: in case of similar mating, the population member that is the most similar to the first individual is chosen as the mate; in case of dissimilar mating, the most dissimilar individual is chosen. That is, the selection-for-mating step of a simple GA (Mitchell, 1996) is modified as:

During each mating event, a fitness-proportionate selection is run to pick out the first individual. Then the Hamming distances of all population members to the first individual are calculated. The actual mate of the first individual is chosen according to the following two different schemes:

**Maximum Similar Mating (MSM):** The population member whose Hamming distance is the smallest is selected for mating.

**Maximum Dissimilar Mating (MDM):** The population member whose Hamming distance is the largest is selected for mating.

Notice that in the mate selection schemes above if several members are of the same maximum (or minimum) Hamming distance to the first individual, then one of them is randomly selected. The computational cost involved in a generation is  $O(N^2)$  similarity comparisons, where  $N$  is the population size.

These two extreme cases simplify the analysis of mate selection based on the

Schema Theorem (Holland, 1975). The next subsection presents such a theoretical analysis.

### 2.1.1 The Schema Theorem Analysis

The investigation on the increase and decrease of the number of schema instances over generations makes clear how GAs process sub-solutions to improve the search quality. Holland's schema theorem (1975), which describes such a dynamic behavior, is:

$$E(m(H, t + 1)) \geq m(H, t) \cdot (f(H, t)/\bar{f}(t)) \cdot [1 - p_c(\frac{d(H)}{l-1})] \cdot [(1 - p_m)^{o(H)}], \quad (1)$$

where  $E(m(H, t))$  represents the expected number of instances of schema  $H$  present in the population at generation  $t$ ;  $m(H, t) \cdot (f(H, t)/\bar{f}(t))$  is the expected times of  $H$  being selected;  $1 - p_c(\frac{d(H)}{l-1})$  is the minimum probability that  $H$  survives crossover, and  $(1 - p_m)^{o(H)}$  is the probability that  $H$  survives mutation.

This theorem is often interpreted as implying that short, low-order, above-average (in fitness) schemata grow exponentially over time, while below-average schemata decay at a similar rate. Therefore, the simple GA, prior to significant convergence, allocates an exponentially increasing number of trials to promising schemata or building blocks.

In the mechanism of simple GAs (Mitchell, 1996), selection is done *with replacement*, meaning that the same individual can be selected more than once to become a mate. This includes the likelihood of mating with the exactly same copy of oneself. Since in case of the maximum similar mating an individual chooses the population member who is the most similar to itself, this would guarantee that the first individual always chooses its identical copy for mating. Thus the probability that schema  $H$  survives crossover is 1.

Let  $E_{MSM}(m(H, t))$  denote the expected number of instances of  $H$  at the  $t$ th generation based on the maximum similar mating. Then the schema theorem becomes:

$$E_{MSM}(m(H, t + 1)) = m(H, t) \cdot (f(H, t)/\bar{f}(t)) \cdot (1 - p_m)^{o(H)}.$$

On the other hand, in case of the maximum dissimilar mating an individual chooses the population member whose Hamming distance is the largest. This would give the crossover-surviving probability of schema  $H$  a larger likelihood to approach the lower bound  $1 - p_c(\frac{d(H)}{l-1})$ . Since Equation 1 still describes schema  $H$ 's dynamic behavior, let  $E_{MDM}(m(H, t))$  represent the expected number of instances of  $H$  at the  $t$ th generation based on the maximum dissimilar mating. Then the corresponding schema theorem is:

$$E_{MDM}(m(H, t + 1)) \geq m(H, t) \cdot (f(H, t)/\bar{f}(t)) \cdot [1 - p_c(\frac{d(H)}{l-1})] \cdot [(1 - p_m)^{o(H)}].$$

The ratio of the expected growth rate of schema  $H$  for these two mate selection schemes is given by

$$1 \leq \frac{E_{MSM}(m(H, t + 1))}{E_{MDM}(m(H, t + 1))} \leq \frac{1}{1 - p_c(\frac{d(H)}{l-1})}. \quad (2)$$

Equation 2 thus shows that the expected schema-growth rate for the maximum similar mating is greater than that for the maximum dissimilar mating.



Table 1: Small royal road function  $S_1$ .

$s_1 = 11111*****;$	$c_1 = 10$
$s_2 = *****1111*****;$	$c_2 = 10$
$s_3 = *****1111*****;$	$c_3 = 10$
$s_4 = *****1111*****;$	$c_4 = 10$

## 2.2 Empirical Results

The two cases proposed thus far facilitate the theoretical analysis. Although they represent two opposite extremes, it turns out that such an analysis can be applied to the investigation for some well known phenomena in the context of GA—hitchhiking and the founder effect, which have been identified as important factors that affect the GA’s search power.

### 2.2.1 Hitchhiking

In population genetics “hitchhiking” is a well-known phenomenon that occurs when some newly discovered allele (or sets of alleles) offers great fitness advantages. As that allele spreads quickly through the population, the closely linked alleles (though they may make no contribution to the fitness) could propagate to the next generation by hitchhiking on that allele. The rapid occupancy of those non-relevant alleles thus greatly reduces exploration of alternatives at those loci. They either drown out other already-discovered alleles that are advantageous, or leave no room for not-yet-discovered beneficial alleles.

In GA research, hitchhiking has been identified as a major problem that limits implicit parallelism by reducing the sampling frequency of various building blocks (Das and Whitley, 1991; Forrest and Mitchell, 1993). Forrest and Mitchell found that if some intermediate stepping stones are much fitter than the primitive components, hitchhiking generates more severe problems that greatly hamper the discovery of some necessary schemata.

To see the hitchhiking phenomenon in the context of GA, let us consider an example function—a small “Royal Road (RR)” function (Forrest and Mitchell, 1993), in which four consecutive building blocks of five ones each are defined. Table 1 is the schematic of this function.

This function involves a set of schemata  $S = \{s_1, \dots, s_4\}$  and the fitness of a bit string  $x$  is defined as

$$F(x) = \sum_{s \in S} c_s \sigma_s(x),$$

where each  $c_s$  is a value assigned to the schema  $s$  as defined in the table;  $\sigma_s(x)$  is defined as 1 if  $x$  is an instance of  $s$  and 0 otherwise. In this function, the fitness of the global optimum string (20 1’s) is  $10 \times 4 = 40$ .

This small Royal Road function is selected as a testbed because it belongs to a class of building-block-based functions, in which improvements in the RR domain depend entirely on the discovery and exploitation of building blocks. This serves as an idealized testbed for us to observe (1) how mate selection facilitates distinguishing individuals that carry necessary building blocks for further improvements, and then (2) how crossover brings these building blocks residing on separate strings into combination on a single string. (The second goal has been extensively investigated by Forrest et

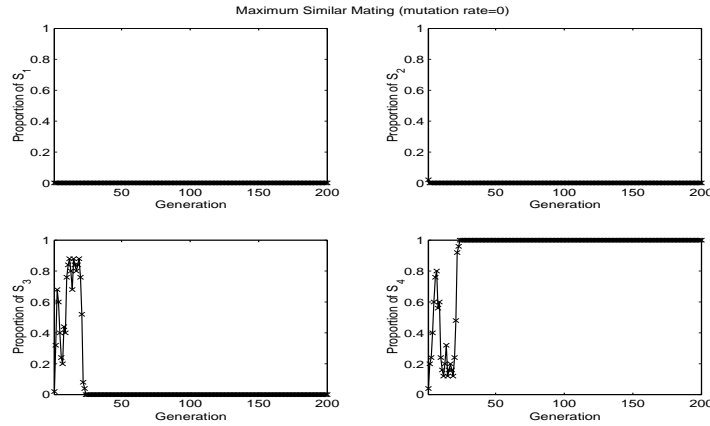


Figure 1: Schema dynamics for observing hitchhiking (Maximum Similar Mating).

al. (1993). This paper focuses on the study of the first goal.)

One can observe hitchhiking directly by plotting the densities (percentage of the population that are instances) of the relevant schemata over time for the GAs. The experiments performed are based on one-point crossover rate 1,<sup>1</sup> and population size 50 over 200 generations.

To give hitchhiking its easiest test, I turn off the mutation operator, since mutation may destroy hitchhiker(s). Figure 1 is a typical run that illustrates density dynamics for the GA based on the maximum similar mating. Schemata  $s_3$  and  $s_4$  are found in the first several generations. Then these two schemata compete against each other before generation 20. Thereafter, schema  $s_4$  takes over the population and quickly spreads through the population. In the mean time, schema  $s_3$  was rapidly suppressed by the hitchhikers adjacent to  $s_4$ . A closer examination shows that these hitchhikers are 00110. Since mutation is turned off and mating partners are simply the copies of individuals that select them, the population is quickly filled with the clones of the individuals that carry  $s_4$  and its hitchhikers, and finally converges to a single genotype.

Next we examine the GA runs based on the maximum dissimilar mating, with the same random seed. Figure 2 displays the density plots for a typical GA run with zero mutation rates. The appearances and disappearances of blips are more dramatic than Figure 1. Unlike Figure 1,  $s_4$  has no apparent edge over  $s_3$ : the densities of the two schemata never exceed 0.6, leaving enough space for adjacent schemata to reproduce in the population and stay alive for a longer time.

The difference between the experimental results obtained for these two mating schemes is a consequence of different schema growth rates. As shown in the preceding subsection, the schema growth rate for the maximum similar mating is greater than that for the maximum dissimilar mating. As a result, once a schema is discovered, the maximum similar mating guarantees that, with a larger likelihood, instances of that schema take over the entire population in a shorter time than the maximum dissimilar mating. This often entails non-relevant hitchhiking alleles that impede further exploration of alternatives at the hitchhiking loci. Then the density of one or more of the

<sup>1</sup>Crossover rate 1 was used in the hope of increasing the improvement rate if two mating partners fit each other. E.g., if crossover rate is 1, then the mating between 11110000 and 00001111 (see the example in Section 1.1) immediately generates the optimum string (11111111) given that the crossover point is exactly at the middle of the strings.

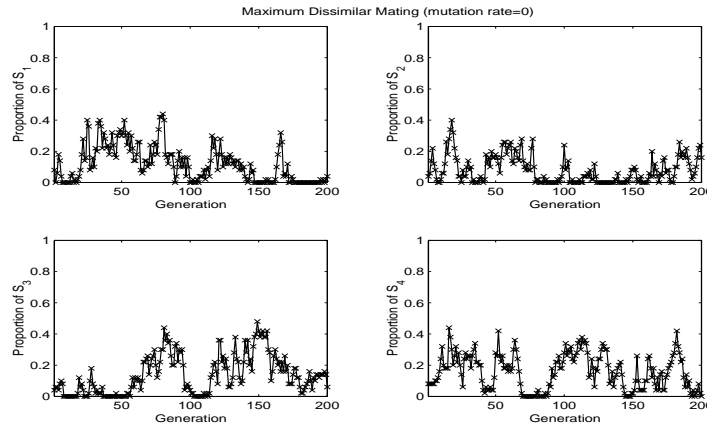


Figure 2: Schema dynamics for observing hitchhiking (Maximum Dissimilar Mating).

disjoint schemata is seen to decline significantly. The greater convergence rate resulting from the similar mating scheme therefore can negate progress that the population has made towards good schemata that overlap the hitchhiking bits.

The most likely positions for hitchhikers are those close to the defined positions of good schemata, since they are less likely to be separated from those defined positions under crossover (Forrest and Mitchell, 1993). This implies that the defining bits of those schemata in the highly fit strings, along with the *nearby* bits, quickly come to occupy most of the population. In other words, the nearby bits “hitchhiking” to prominent schemata partially or totally suppress the appearance of disjoint building blocks, leaving few variants at those loci. This leads to the loss of diversity in the vicinity of the better schemata. As a consequence, some of the necessary building blocks for crossover to combine to gain performance advancements are lost or unlikely to be discovered, and the GA’s search power is greatly hampered.

### 2.2.2 Founder Effect

In GA research, a much more important constraint on exploration than hitchhiking is the founder effect (Holland, 2000). In presence of incompatible schemata, the first discovered of the incompatible schemata comes to occupy a large portion of the population, and constrain future evolutionary progress. Consequently, the founder schema effectively precludes the testing of the other incompatible schema. Further improvements stem from the founder, making it progressively less likely that the other schema will influence the search process.

The fitness landscape of the simple royal road function used in the preceding subsection consists of four consecutive building blocks. The combination of these building blocks constructs the only path for the GA to improve the search. Consider a variant of the simple royal road function that consists of incompatible schemata as shown in Table 2.

This function involves mutually exclusive alleles at each schema, and there are 16 ( $2^4$ ) alternatives for search to improve solution quality. The fitness of the global optimum (20 1’s) is 40, and that of the other local optimum (20 0’s) is 20.

The incompatibility of schemata on this function is designed for testing the founder effect—I would expect each 5-bit block to be soon occupied by a founder

Table 2: Incompatible small royal road function  $IS_1$ .

$s_1 = 11111*****;$	$c_1 = 10$
$s_2 = 00000*****;$	$c_2 = 5$
$s_3 = *****11111*****;$	$c_3 = 10$
$s_4 = *****00000*****;$	$c_4 = 5$
$s_5 = *****11111*****;$	$c_5 = 10$
$s_6 = *****00000*****;$	$c_6 = 5$
$s_7 = *****11111*****;$	$c_7 = 10$
$s_8 = *****00000*****;$	$c_8 = 5$

schema, unless the founder schema is destroyed by crossover or mutation.

Again one can examine the founder effect directly by plotting the density of each schema over generations. The experiments performed are based on one-point crossover rate 1, and population size 50 over 200 generations.

Analogous to the discussion for the hitchhiking problem, mutation is turned off in order to give the founder effect its clearest examination. Figure 3 is a typical run that displays the schema density dynamics for the maximum similar mating GA. Schemata  $s_1$  and  $s_4$  are discovered at the first generation and quickly spread through the population. A closer examination shows that each of  $s_1$  and  $s_4$  has an instance at the first generation. Then because maximally similar individuals are always chosen for mating, which in turns enhances the degree of premature convergence, the instances of  $s_1$  and  $s_4$  quickly take over the whole population. Due to the incompatibility,  $s_2$  and  $s_3$  are prohibited from being tested. In the mean time, along with the growth of schema  $s_4$  hitchhikers 10110 rapidly propagate to the whole population and block the testing of schema  $s_5$  or  $s_6$  (adjacent to  $s_4$  on the right).

These results demonstrate that for this typical run  $s_1$  and  $s_4$  are the first discovered of the respective, incompatible schemata which rapidly found their dynasties. As a consequence, further improvements are constrained by  $s_1$  and  $s_4$ , and the GA can only devote itself to 4 possible alternatives for search improvements. (Due to sampling errors, either of each pair of incompatible schemata is likely to dominate the population. Therefore, for other runs different schemata may take over the whole population.)

For the maximum dissimilar mating, the density plots for a typical GA run with mutation being turned off is displayed in Figure 4. Unlike Figure 3, no schema effectively founds a dynasty. As can be seen, the maximum density of each schema is around 0.5, making it less likely that either of incompatible schemata precludes the testing of other incompatible schema (note that, due to sampling error,  $s_2$ ,  $s_4$ , and  $s_5$  still do not show up in the population in later generations). Since the maximum dissimilar mating allows individuals to exchange genetic material with dissimilar mates, they can recruit more distinct alleles to escape from the founder effect and hitchhiking.

### 2.2.3 Rate of Improvement and Creation of Lethal Hybrids

How rapid is evolutionary change, and what determines the rates, patterns, and causes of change, or lack thereof? Answers to these questions can tell us much about the evolutionary process. The study of evolutionary rate in the GA context usually involves defining a performance measure that embodies the idea of rate of improvement, so that its change over time can be monitored for investigation. In this paper, the performance

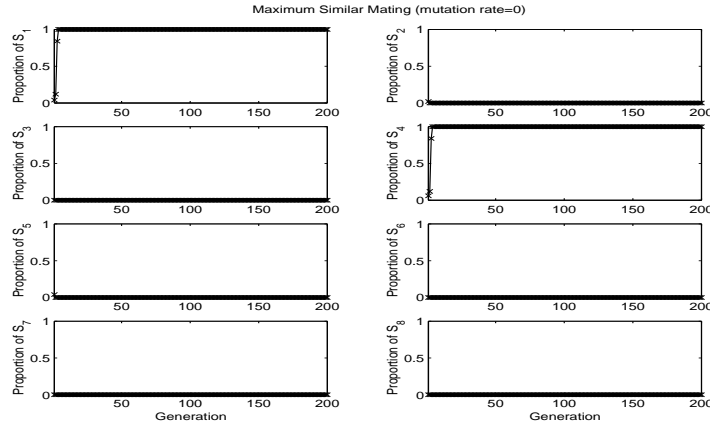


Figure 3: Schema dynamics for observing the founder effect (Maximum Similar Mating).

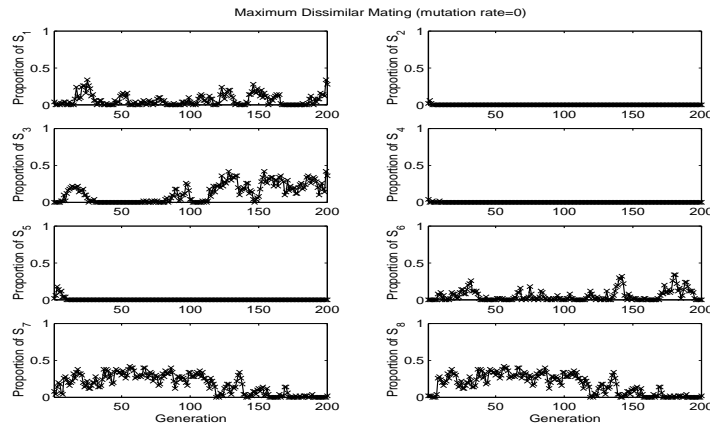


Figure 4: Schema dynamics for observing the founder effect (Maximum Dissimilar Mating).

metric employed is quite traditional—“best-so-far” curves that plot the fitness of the best individual that has been seen thus far by generation  $n$ .

Since the two mating schemes discussed so far are two extreme cases where only the maximally similar or dissimilar population member is chosen as the mate, I introduce another two intermediate mate selection schemes before studying the rate of improvement. That is, during each mating event, the first individual is still picked by fitness-proportionate selection, and the Hamming distances of all population members to this individual are calculated. Then the mate for the first individual is chosen according to the following two schemes:

**Proportional Similar Mating (PSM):** The probabilities of population members being selected are *inversely* proportional to their Hamming distances.

**Proportional Dissimilar Mating (PDM):** The probabilities of population members being selected are proportional to their Hamming distances.

Notice that since individuals' Hamming distances may be zero, to avoid dividing by zero in the proportional similar mating I offset all the Hamming distances by 1 in this paper. (Other offset values are possible, of course.)

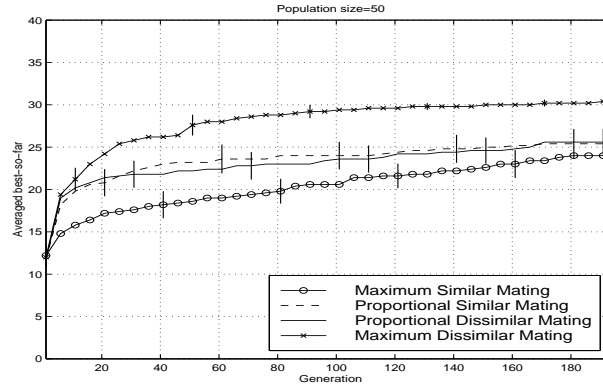


Figure 5: Best-so-far performance on  $S_1$ .

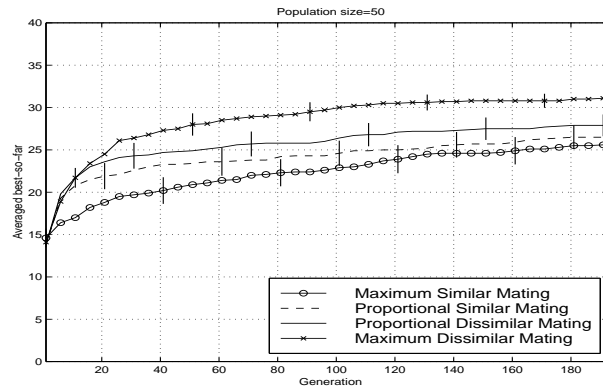


Figure 6: Best-so-far performance on  $IS_1$ .

The experiments conducted here are based on one-point crossover rate 1, mutation rate 0.005, and population size 50 over 50 runs. Figure 5 and 6 display the averaged best-so-far curves on  $S_1$  and  $IS_1$ , respectively.<sup>2</sup> One can see that dissimilar matings tend to provide advantage to the GA's search power, especially the maximum dissimilar mating scheme.

In (Huang, 2002a) it is shown that, by suppressing hitchhiking and the founder effect, the maximum dissimilar mating retains more genetic variation in the population. By further exploring the search space the GA thus yields a better best-so-far search progress, yet at the expense of the mean population fitness. Namely, although the maximum dissimilar mating GA is engaged in searching and constructing better best-so-far solutions, this mate selection would seem to create more lethal hybrids that degrade the population's fitness.

<sup>2</sup>The vertical bars overlaying the metric curves throughout this paper represent the 95-percent confidence intervals calculated from Student's  $t$ -statistic (Miller, 1986).

The following example supports this claim. Suppose the population is composed of bit-strings of length 4 and the relevant building blocks that are able to contribute are 11\*\* and \*\*11 (the optimal string is thus 1111). Consider lethal offspring that are easily removed from the population, say individuals of fitness 0. Then, for example, given a string  $X$ , 1100, and two candidate mating partners,  $Y_1$  (1100) and  $Y_2$  (0011), the maximum similar mating requires  $X$  to select  $Y_1$  as the mate. Since  $X$  and  $Y_1$  are identical, their mating generates neither the optimum nor lethal hybrids. In case of the maximum dissimilar mating, however,  $Y_2$  will be chosen as the mate and there exists probability  $\frac{1}{3}$  for generating the optimum and  $\frac{2}{3}$  for lethal offspring after crossover.

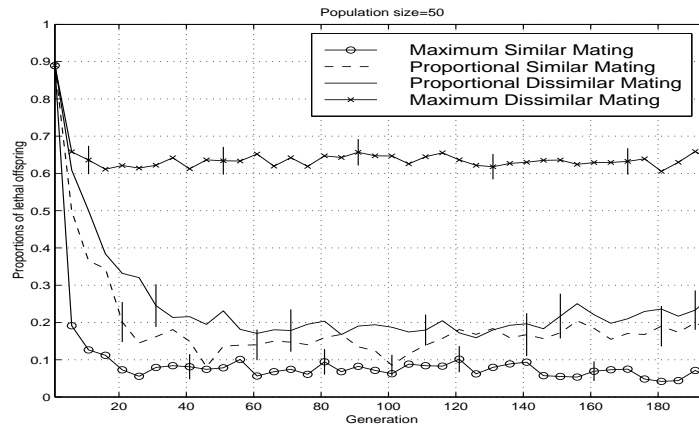


Figure 7: Proportions of lethal offspring on  $S_1$ .

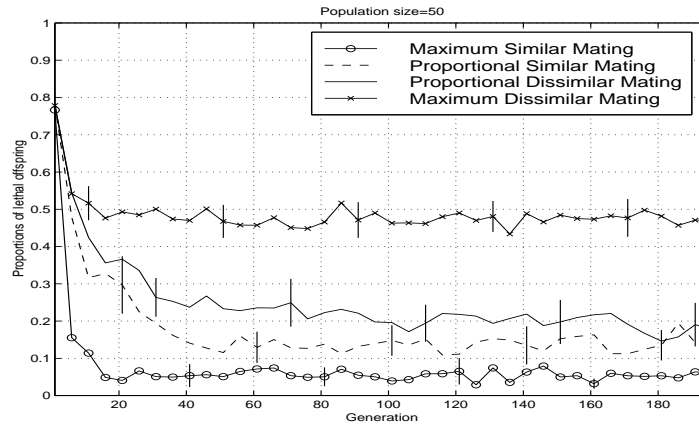


Figure 8: Proportions of lethal offspring on  $IS_1$ .

The empirical results in Figure 7 (based on the test function  $S_1$ ) validate the claim made above, from which one can see the significant proportions (between 0.6 and 0.7) of lethal offspring generated from the maximum dissimilar mating. This indicates that for the maximum dissimilar mating 30 to 35 individuals (population size 50) in each generation are lethal and rapidly disappear from the population. Similar matings, on the other hand, prevent the creation of lethal offspring. In particular, matings between

maximally similar individuals preserve the proportions of lethal offspring at the level lower than 0.1 after a few generations, meaning that on average less than 5 strings are of fitness value 0 (in fact, these lethal strings are generated from mutation, because crossover has minimal effects on the maximum similar mating scheme). Figure 8 illustrates similar results for the test function  $IS_1$ .

### 2.3 Summary

In this section the framework proposed for investigating mate selection allows us to conduct an analysis to deepen our understanding of GAs. I first started with mating preference for the maximally similar or dissimilar individuals, and showed that, in the context of GA, hitchhiking and the founder effect can be explained in terms of the schema analyses, and these two phenomena can be suppressed by dissimilar mating choices. The advantage of mating with dissimilar individuals is reflected by the corresponding GA's improved best-so-far performance.

Afterwards I introduced more complicated means for calculating individuals' probabilities of being selected via combining similarity test and fitness-proportionate selection. Allowing population members to discriminate candidate mates provides another source of selection pressure, in addition to the selection pressure arising from the environment.

In short, the analyses reveal that individuals with more distinct characteristics (in terms of the Hamming distance) facilitate the search for a single, better solution. The problem it may cause is lethal hybrids: matings between dissimilar individuals give crossover more opportunity to disrupt existing building blocks, leading to decrease of the mean population fitness. (See (Huang, 2002a) for a more detailed discussion.) Therefore, to avoid producing lethal offspring may require that matings occur between individuals of similar characteristics.

In the next section, I present a Markov chain analysis for further study of these mate selection strategies.

## 3 Markov Chain Analysis

Since a "state" of simple genetic algorithms can be defined by a particular population, and the composition of the population at the next step is entirely determined by the present population, it has been quite natural to model simple GAs as Markov chains. For example, there are several Markov models that were derived assuming infinite population and involve characterizing steady state behavior (Suzuki, 1993; Rudolph, 1994). By contrast, the model developed by Nix and Vose (1992) was based on finite population size. De Jong, Spears, and Gordon (1994) applied this finite-population model to investigate GA-based function optimization (GAFO), in which they performed a transient Markov chain analysis to calculate the mean waiting times. In this section, I intend to further study these mate selection schemes using a Markov chain model, based on the Nix and Vose model and the GAFO theory of De Jong et. al. Although such models quickly become unwieldy with increasing population size or string length, they provide important insights into how mate selection plays a crucial role in GA's search power, and thus serve as guidelines for studying more realistic problems.

In the following subsection, the Nix and Vose Markov model is briefly summarized. Then in Subsection 3.2 I show how the mate selection schemes are incorporated in this model, and use them for simple visualization analysis. Subsection 3.3 is a brief summary of the GAFO theory of De Jong et. al., following some computational explorations for understanding effects of mate selection. The final subsection presents



Table 3: The  $Z$  matrix ( $n=2$  and  $l=2$ ).

State	00	01	10	11
P1	0	0	0	2
P2	0	0	1	1
P3	0	0	2	0
P4	0	1	0	1
P5	0	1	1	0
P6	0	2	0	0
P7	1	0	0	1
P8	1	0	1	0
P9	1	1	0	0
P10	2	0	0	0

relevant discussions.

### 3.1 The Nix And Vose Markov Model

Nix and Vose (1992) introduced a Markov chain model that is intended to represent a simple, finite-population GA based on a standard binary representation, fitness-proportional selection scheme, standard mutation, and one-point crossover operators. The only difference between their model and standard GAs (Mitchell, 1996) is that only one offspring from each crossover survives. As a result, for population size  $n$ , a total of  $n$  recombination events take place. (This simplifies parts of the formalization.)

The simple GA model they considered moves from one generation to the next as follows:

- (1) Obtain two parents by proportional selection.
- (2) Mutate (mutation implies change) the parents with rate  $\mu$ .
- (3) Produce the (mutated) parents' child by one-point crossover with rate  $\chi$ .
- (4) Put one child into the next generation.
- (5) If the next generation contains less than the population size, go to step 1.

In their model, the search space is the set of all length  $l$  strings and  $r = 2^l$  is the total number of possible strings. If  $n$  is the population size, then the number of possible populations,  $N$ , corresponding to the number of possible states is:

$$N = \binom{n + 2^l - 1}{2^l - 1}.$$

The possible populations are described by the matrix  $Z$ , which is an  $N \times r$  matrix.<sup>3</sup> The  $i$ th row  $\phi_i = \langle z_{i,0}, \dots, z_{i,r-1} \rangle$  of  $Z$  is the incidence vector for the  $i$ th population. Thus  $z_{i,y}$  is the number of occurrences of string  $y$  in the  $i$ th population, where  $y$  is the integer representation of the binary string. For example, suppose  $l = 2$  and  $n = 2$ ; then  $r = 4$ ,  $N = 10$  and the  $Z$  matrix is shown in Table 3:

With these definitions, Nix and Vose derived the formula for calculating exact state

<sup>3</sup>For programming convenience I transpose the  $Z$  matrix of Nix and Vose (1992), as indicated by De Jong et al. (1994).

transition probabilities  $Q_{i,j}$ :

$$Q_{i,j} = n! \prod_{y=0}^{r-1} \frac{(M[\frac{F\phi_i}{|F\phi_i|}]_y)^{z_{j,y}}}{z_{j,y}!}, \quad (3)$$

where  $F$  is determined from the fitness function, and  $M$  depends on the mutation and crossover operators.

Significant insights into transient behavior can be obtained computationally by computing and analyzing  $Q^k$  directly. Unfortunately, the size of the  $Q$  matrix for typical GAFO applications is computationally unmanageable since the number of states  $N$  grows rapidly with string length  $l$  and population size  $n$ . For example, a GA with population size 10 and bit-string length 6 has about  $6.2 \times 10^{11}$  states.

### 3.2 Mate Selection In The Nix And Vose Markov Model

The Nix and Vose Markov model consists of two key operators:  $F$  and  $M$ , where  $F$  relates to selection, and  $M$  relates to mutation and crossover. Since the mate selection schemes only affect the GA's selection process, altering the operator  $F$  is the major objective in this section.

In the formal model of Vose and Liepins (1991), each string in the search space is specified by the corresponding integer equivalent between 0 and  $2^l-1$ . The population at generation  $t$  is represented by two vectors,  $\vec{p}(t)$  and  $\vec{s}(t)$ , each of length  $2^l$ , where  $\vec{p}(t)$  specifies the composition of the population at generation  $t$ , and  $\vec{s}(t)$  represents strings' probabilities of being selected. Let  $F$  be a two-dimensional matrix such that  $F_{i,j} = 0$  for  $i \neq j$ , and  $F_{i,i} = f(i)$ , where  $f(i)$  is the fitness of string  $i$ . Then under proportional selection it yields

$$\vec{s}(t) = \frac{F\vec{p}(t)}{\sum_{i=0}^{2^l-1} F_{i,i}p_i(t)}, \quad (4)$$

where  $p_i(t)$  denotes the  $i$ th component of  $\vec{p}(t)$ , which is the proportion of the population at generation  $t$  consisting of string  $i$ . Therefore the  $i$ th component of  $\vec{s}(t)$  is the probability that string  $i$  will be selected as a parent.

For example, if  $l=2$  and the population consists of two copies of 11 and one copy of 00 and 01, then

$$\vec{p}(t) = (0, 0.25, 0.25, 0.5).$$

If the fitness is equal to the number of ones in the string, then

$$\vec{s}(t) = (0, 0.1667, 0.1667, 0.6667).$$

Given these preliminaries, one can compute the expected proportion of string  $k$  at generation  $t+1$ :

$$E(p_k(t+1)) = \sum_{i,j} s_i(t)s_j(t)r_{i,j}(k), \quad (5)$$

where  $r_{i,j}(k)$  is the probability that string  $k$  will be produced by a recombination event between string  $i$  and string  $j$ , given that  $i$  and  $j$  are selected to mate.

Then Vose and Liepins derived the operator  $M$  to encapsulate the calculations regarding crossover and mutation that account for  $r_{i,j}(k)$ . Finally, Nix and Vose (1993) followed these results to construct the finite-population model as described in Equation 3.

Table 4: Procedure for calculating the second individual's probability of being selected in Equation 5 for the maximum similar mating.

1. Let  $F_{i,j}=0$  for all  $i$  and  $j$ .
2. For  $i$  from 1 to  $r$ , do:
  - a. For  $j$  from 1 to  $r$ , compute the Hamming distance of string  $j$  to string  $i$ , if  $z_{m,j} \neq 0$ , where  $z_{m,j}$  is the number of occurrences of string  $j$  in the  $m$ th population.
  - b.  $F_{j,j}=1$ , if string  $j$  is the only one that possesses the minimum Hamming distance; or  $F_{j,j} = \frac{1}{u}$ , if  $u$  is the number of strings whose Hamming distances are the minimum.

To understand how the mate selection schemes are integrated with the Nix and Vose Markov model, let us take a closer look at Equation 5, in which both the individuals selected for mating—string  $i$  and string  $j$ —are based on fitness-proportionate selection. Recall that the maximum similar mating scheme picks up the first individual, which plays the role of string  $i$  in Equation 5, by fitness-proportionate selection. Then the population member who is the most similar to the first one is chosen as the second individual, which appears as string  $j$  in Equation 5. Hence the way of calculating  $s_j(t)$  in Equation 5 is the only part that needs to be reconsidered.

Since the maximally similar individual must be chosen as the second string, it turns out that the fitness matrix  $F$  needs to be re-defined so that only the string that is maximally similar to the first string can possess non-zero fitness values. If several strings are of the same minimum Hamming distance to the first string, these strings are assigned the same fitness values.

Plugging the new fitness matrix  $F$  into Equation 4 yields the new selection probability  $s_j(t)$  of string  $j$  for Equation 5. The resulting Markov model is exactly the GA model with the maximum similar mating scheme.

This implementation is summarized in Table 4.

As for the maximum dissimilar mating, the first string will select as its mate the string whose Hamming distance is the maximum. Thus the implementation of the maximum dissimilar mating remains the same as that of the maximum similar mating, except that “minimum” is replaced with “maximum” in Table 4.

We can proceed to derive the Markov model for the proportional similar mating in the same way. Recall that in the proportional similar mating, the first string is picked by fitness-proportionate selection, and the probabilities of strings being selected as the first string's mate are *inversely* proportional to their Hamming distances. As discussed previously, I offset all the Hamming distances by 1. The implementation is shown in Table 5.

As for the proportional dissimilar mating, since the probabilities of population members being selected as the mate are proportional to their Hamming distances, the corresponding implementation can be obtained by replacing “ $1/(D_j + 1)$ ” in Table 5 with “ $D_j$ ”.

Table 5: Procedure for calculating the second individual's probability of being selected in Equation 5 for the proportional similar mating.

1. Let  $F_{i,j}=0$  for all  $i$  and  $j$ .
2. For  $i$  from 1 to  $r$ , do:
  - a. For  $j$  from 1 to  $r$ , compute the Hamming distance of string  $j$  to string  $i$ , if  $z_{m,j} \neq 0$ , where  $z_{m,j}$  is the number of occurrences of string  $j$  in the  $m$ th population.
  - b.  $F_{j,j} = \frac{1}{D_j+1}$ , where  $D_j$  represents the Hamming distance of string  $j$ .

### 3.2.1 Visualizing Markov Models And Diversity

Having modified the Markov model for the mate selection schemes, we can compute the state transition probability matrix for each case. Let  $Q_{MSM}$ ,  $Q_{PSM}$ ,  $Q_{PDM}$ , and  $Q_{MDM}$  correspond to the state transition probability matrices for the maximum similar mating, the proportional similar mating, the proportional dissimilar mating, and the maximum dissimilar mating, respectively.

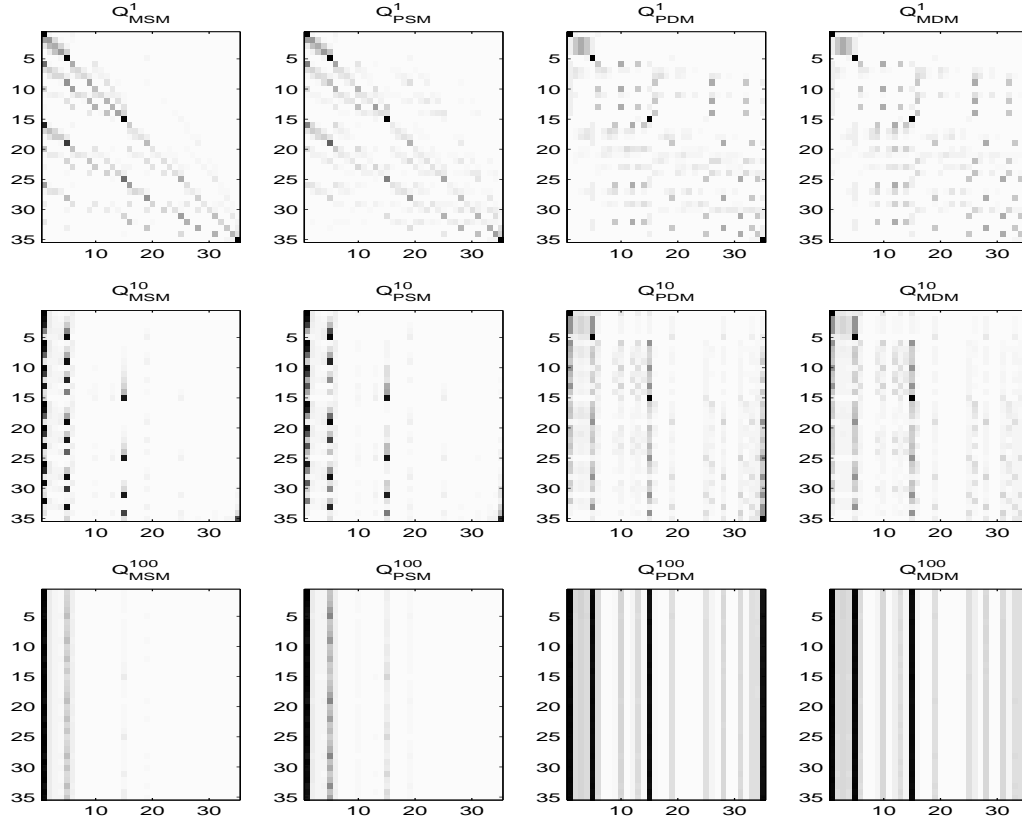
Visualizing  $Q^k$  provides us with the insight regarding the effects that different mate selection schemes have on the state transition probability matrix  $Q$ . (See (De Jong, Spears, and Gordon, 1994) for additional evidence concerning the usefulness of this approach.) This can be implemented by plotting  $Q^k$  as an image, the gray level of coordinate  $(i,j)$  reflecting the probability that the GA will move from state  $i$  to state  $j$  in  $k$  steps. The density of blackness at each  $(i,j)$  indicates the corresponding  $k$ -step transition probability, where the darker a point, the higher the corresponding probability.

As an example, I use a fitness function with  $f(y) = \text{integer}(y) + 1$  (De Jong, Spears, and Gordon, 1994), where  $\text{integer}(y)$  returns the integer equivalent of the bit string  $y$ . The string length is 2, population size is 4 (thus the number of total states is 35), crossover rate is 1, and mutation rate is 0.01.

Figure 9 shows the results for various  $Q_m^k$ , where  $m$  represents MSM, PSM, PDM and MDM. Let us first inspect the case for one-step transition, i.e.,  $k=1$ . A clear, visible diagonal line from  $Q_{MSM}^1$  indicates that significant changes in the population in one generation are very unlikely. As we scan the images from left to right, we see that this situation can be altered by gradually allowing matings to take place between relatively dissimilar individuals, generating more diffuse state transition probability matrices. The net result is that dissimilarity-based matings can make larger changes more easily.

As time proceeds, we see that the changes in the probability distribution are already evident in  $Q_m^{10}$  and even more evident in  $Q_m^{100}$ . The emerging vertical lines correspond to the particular populations at which the steady state distribution will accumulate most of its probability mass (i.e., the populations most likely to be observed when the GA settles into its dynamic equilibrium).

We can use the Markov model to further examine whether or not dissimilarity-based mate selection schemes will maintain larger population diversity. Recall (from the preceding subsection) the matrix  $Z$  represents possible populations, which is an  $N \times r$  matrix. The  $i$ th row  $\phi_i = \langle z_{i,0}, \dots, z_{i,r-1} \rangle$  of  $Z$  is the incidence vector for the  $i$ th

Figure 9: Visualization of  $Q^k$ .

population, and  $z_{i,y}$  is the number of occurrences of string  $y$  in the  $i$ th population.

Let  $b$  denote the row vector of strings' binary representations, in which the  $i$ th component corresponds to string  $i$ 's binary representation ( $i = 0, \dots, r - 1$ ). Then  $b_{i,j}$  represents the bit value of string  $i$ 's locus  $j$  ( $j = 1, \dots, l$ ). For instance, if binary strings are of length 2, then  $b = \langle 00, 01, 10, 11 \rangle$ , and  $b_{0,1} = 0$ ,  $b_{1,2} = 1$ ,  $b_{3,1} = 1$  and so forth. To compute the proportion of 1's at each locus in the  $i$ th population, we need to extend the definition of vector inner product to include products of scalars and bit strings.

For example,  $\phi_2 = \langle 0, 0, 1, 1 \rangle$  represents state P2 of Table 3, and its inner product with  $b = \langle 00, 01, 10, 11 \rangle$  is:

$$\begin{aligned} \sum_{i=0}^3 z_{2,i} \cdot b_i &= 0 * (00) + 0 * (01) + 1 * (10) + 1 * (11) \\ &= (10) + (11). \end{aligned}$$

The result above shows that the number of 1's at the first locus (the right bit of strings) is 1, and that at the second locus (the left bit of strings) is 2. Dividing the number of 1's at each locus by population size 2 yields:

$$p_1 = 0.5 \quad \text{and} \quad p_2 = 1,$$

where  $p_j$  represents the proportion of 1s at locus  $j$ .

Thus the inner product of  $\phi_i$  and  $b$ , dividing by population size  $n$ , represents the proportion of 1's at each locus in the  $i$ th population.

Let  $p_{i,j}$  be the proportion of 1s at locus  $j$  in the  $i$ th population. To measure diversity at the  $j$ th locus, a simple bitwise diversity metric can be defined as follows (Mahfoud, 1995):

$$D_{i,j} = 1 - 2|0.5 - p_{i,j}|. \quad (6)$$

One can use the average of the bitwise diversity over all loci as a combined allelic diversity measure for the population. Let  $DIV_i$  represent the allelic diversity for the  $i$ th population. Then we have the allelic diversity measure for the  $i$ th population:

$$DIV_i = \frac{\sum_{j=1}^l D_{i,j}}{l}. \quad (7)$$

$DIV_i$  has a value of 1 when the proportion of 1s at each locus is 0.5 and 0 when all of the loci are fixed to either 0 or 1. Effectively it measures how close the allele frequencies are to a random population (1 being closest).

Let  $DIV$  denote the column vector of populations' averaged allelic diversity. Then the expected allelic diversity at generation  $k$  is:

$$E(DIV) = (iniP \cdot Q^k) \cdot DIV, \quad (8)$$

where  $iniP$  is a row vector whose  $i$ th component represents the probability of the GA being in state  $i$  at generation 0 (the initial generation), and thus  $iniP \cdot Q^k$  is states' probability distribution at generation  $k$ . In this paper, I consider randomly initialized GAs. The probability of a GA being in state  $i$  at time 0, denoted as  $P(X_0 = i)$ , is (see (De Jong, Spears, and Gordon, 1994)):

$$P(X_0 = i) = \frac{n!}{z_{i,0}! \cdots z_{i,r-1}!} \left(\frac{1}{r}\right)^n.$$

By Equation 8, we can compute the exact average allelic diversity for any cases. For illustrations, I still use the fitness function with  $f(y) = \text{integer}(y) + 1$ . Figure 10 displays the results for string length 2 and several different population sizes ( $n = 3, \dots, 8$ ), based on crossover rate 1 and mutation rate 0.01. The results show that the averaged allelic diversity of dissimilarity-based mate preferences is larger than that of similarity-based mate preferences. (This agrees with our intuition that dissimilar mating can generate more diverse populations than similar mating.) In particular, one can see that the difference between the dissimilarity and similarity-based mate selection schemes enlarges as population size increases.

### 3.3 The GAFO Theory

The preceding section discussed GA's search power in terms of the best-so-far performance. Such a performance metric is a form that a GA practitioner who is interested in optimization would generally care about. For simple test functions the optimum is easy to locate and one may use the expected time (generations) to first encounter the optimum as a metric to compare different GAs' search power. In this subsection I use the GAFO framework developed by De Jong et al. (1994) to examine the effects of mating choices on GA's performance.

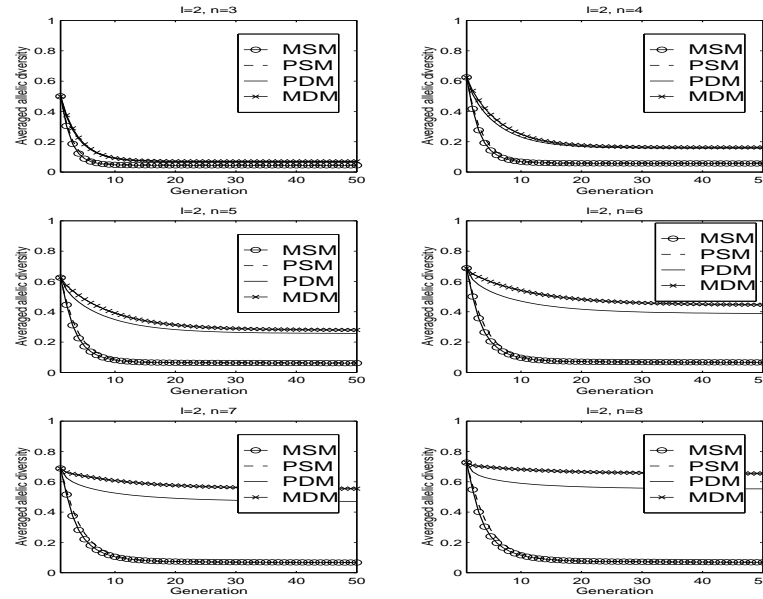


Figure 10: Averaged allelic diversity.

### 3.3.1 Expected Waiting Time Analysis

De Jong et al. (1994) extended the Nix and Vose Markov chain analysis to provide answers for the expected waiting time until an event of interest is first observed. The observation is that the state transition probability matrix  $Q$  can be used to compute “mean first passage times” for going from state  $i$  to state  $j$ . If one is interested in knowing how long the GA would have to run on average before first reaching a set  $J$  of states, given that the process is currently in state  $i$ , then the expected waiting time is:

$$EWT(J) = \sum_{i \notin J} P(X_0 = i) m_{i,J},$$

where  $m_{i,J}$  denotes the mean first passage time from state  $i$  to any of the states in set  $J$ , and  $i$  is not in  $J$ ;  $m_{i,J}$  can be computed from the system of simultaneous equations

$$m_{i,J} = \sum_{j \in J} Q_{i,j} + \sum_{k \notin J} Q_{i,k} (1 + m_{k,J}).$$

If one defines  $J$  to be the set of states containing at least one copy of the optimum string, then  $EWT(J)$  is the expected number of generations until the optimum is first encountered.

Given the summary of De Jong et al.’s work, I present some preliminary results to study effects of various factors on the EWT analysis. Characterizing these effects provides useful insights and make predictions about how to improve GA’s performance in the context of function optimization.

### 3.3.2 Interacting Effects of Mate Selection and Mutation

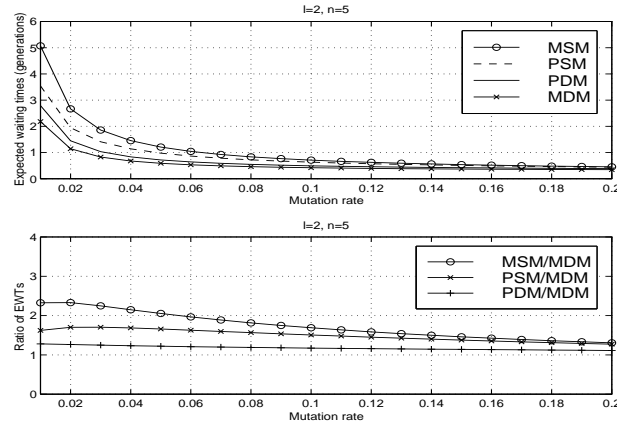


Figure 11: Interacting effects of mate selection and mutation.

As an example, I again use the test function  $f(y) = \text{integer}(y) + 1$ . Due to the computational limitation of the Nix and Vose model, I use the simplest possible case, string length 2, to proceed the investigation. The optimum string is 11. The population size  $n$  used is 5 and crossover rate is 1. For this case the first goal is to investigate the interacting effects of mate selection and mutation on the GA's EWTs to the optimum.

Figure 11 displays the results obtained for the GAs with MSM, PSM, PDM and MDM, based on mutation rates ranging from 0.01 to 0.2. The top plot is the exact EWTs and the bottom plot corresponds to the ratios of the EWTs for the MSM, PSM, and PDM GAs to that for the MDM GA.

One can see that the maximum dissimilar mating (MDM) generally has the lower expected waiting times than the other three. Note how dissimilar matings become increasingly important as mutation rate decreases (the larger the ratios, the better the relative performance of the maximum dissimilar mating). This shows that, as mutation rate decreases, dissimilar mating would become a dominant factor that brings forth population diversity for further exploration of the search space.

### 3.3.3 Interacting Effects of Mate Selection and Crossover

One can also use these models to analyze the interacting effects of mate selection and crossover on EWTs. Figure 12 shows the results obtained for the four mate selection schemes, based on crossover rates ranging from 0.05 to 1, mutation rate .1, and population size 5. The top plot is for the exact EWTs and the bottom plot displays the ratios of the EWTs for the MSM, PSM, and PDM GAs to that for the MDM GA.

One can see that the maximum dissimilar mating generally has the least expected waiting times than the other three. In particular, the dissimilar mating schemes demonstrate increasingly improved performance as crossover rate increases. This shows that both proper mate selection and crossover must operate together to enhance the power of information exchange in GA's population.

### 3.3.4 Effects of Scaling

One can investigate how difficulties of test functions affect the GA's performance. It is well-known that proportional selection is sensitive to simple linear scaling of the fitness function (De Jong, Spears, and Gordon, 1994). For example, given the test function  $f$  and its variant,  $g = f + 100$ , one would expect that the GA is more incapable of locating



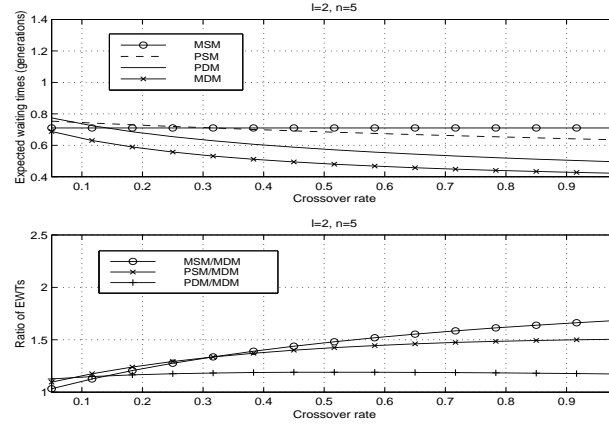


Figure 12: Interacting effects of mate selection and crossover.

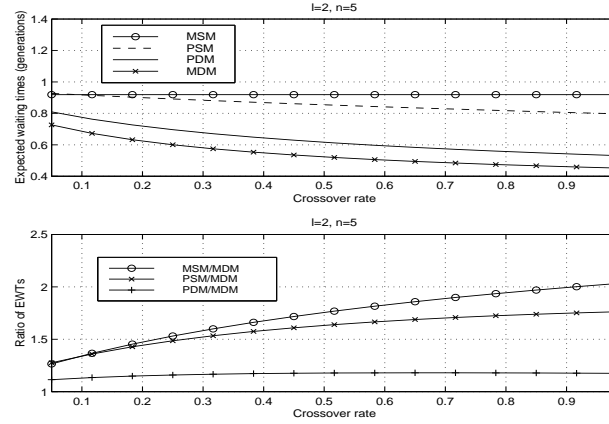


Figure 13: Effects of Scaling.

the optimum of  $g$ , since the fitness of function  $g$ 's optimum is relatively close to other fitness values.

Since MDM and PDM tend to choose dissimilar mating partners, essentially these two dissimilarity-based mating schemes introduce additional effects of magnifying the difference between individuals. (See the simple example illustrated in Section 1.1.) But this is not the case for MSM and PSM. Therefore, if a GA is presented with the test function  $g$ , I predict that the performance difference between dissimilar mating and similar mating is larger than that with the test function  $f$ .

Figure 13 illustrates the EWT analysis on the test function  $g$ . Compared with Figure 12, the performance discrepancy (the ratios of EWTs) between similar mating and dissimilar mating is enlarged, which confirms the hypothesis. These results show that the dissimilarity-based mate selection is beneficial in this more difficult function.

### 3.4 Summary and Discussion

This section presents my initial exploration of transient Markov chain analysis as a theoretical basis for the similarity and dissimilarity-based mate selection schemes. By

computationally characterizing effects of various factors, we gain useful insights concerning the importance of mating choices in GA's search power.

Visualizing the state transition probability matrix  $Q$  is a useful technique to see how the GA population becomes more diverse and diffuse in case of the dissimilar mate selection schemes. I have also used the Markov model to investigate the effects of mate selection on population diversity. The results show that dissimilar matings generate larger population diversity than similar matings.

As mutation rate decreases, dissimilar mating preference in improving the GA's performance in finding the global optimum becomes more important. In addition, the study on the interacting effects of mate selection and crossover shows that both factors working together enhance the GA's search power. I have also demonstrated that the dissimilar mating schemes are beneficial in finding the optimum when test functions are relatively difficult.

A major concern of Markov chain analysis is the scalability of the results. In (Huang, 2002b) the insights discussed above have been applied for further investigation on several more realistic testbeds, with larger population size and string length. The empirical results obtained also show that dissimilar mating is beneficial in improving the GA's search performance.

#### 4 Conclusion and Future Work

In this paper, I introduce simple models to investigate mate selection in the context of GA, where individuals are allowed to search for mates. Such approaches can model interdependent fitnesses of population members and introduce another source of selection pressure. The resulting GA hence forms a more complex system in which individuals' fitnesses depend on both the environment and other population members.

The very essence of good GA design is retention of diversity, furthering exploration, while exploiting building blocks already discovered. Through the systematic investigation, I demonstrate that proper mating preferences are shown to indeed facilitate discriminating individuals in the population, retain genetic diversity, and better utilize building blocks already discovered for exploration of the search space. The advantage of using dissimilarity-based mating preferences is reflected by the corresponding GA's improved best-so-far performance.

Although the results obtained were based on simple fitness landscapes, I have conducted further experiments on more complicated test functions, including more realistic building-block-based and real testbeds. These results will be described in (Huang, 2002b). In particular, several illustrations in that forthcoming paper will also show similar results as those in this paper—i.e., mate selection indeed plays a crucial role in GAs' search power, and selecting dissimilar mates can provide substantial improvement for the GA's best-so-far performance. (The results obtained in (Huang 2002b) are thus encouraging since it means that the framework of mate selection proposed in this paper can be applied to practical problems.)

There are a variety of directions worth exploring. For example, in addition to expected waiting times, the variance of the waiting times is also an important metric that can be derived from the mate selection Markov models. Another direction is to have a systematic investigation on the effects of population size, crossover, mutation, and other factors, either analytically or empirically. All these issues are critical for examining whether mate selection is an important ingredient in GAs.

The investigation so far has been centered around fixed, non-intelligent mate selection schemes. Further research will need to be extended to self-adaptive mate selection,

including the existing tag-added, template-added mechanisms, and others.

Thus far the similarity comparison has focused on the genotypes of population members. In future work, advanced study of mate selection based on individuals' phenotypes will be conducted.

Since this paper focuses on fixed-length, linear chromosomes, one of the other obvious extensions is to variable-length, nonlinear representations, such as those used in Genetic Programming (Koza, 1992). It is clear that proper similarity metrics depend on problem domains and the algorithms used. Therefore, one would need to select suitable similarity metrics for different problem domains.

### Acknowledgments

The author would like to thank John Holland, Rick Riolo for their advice, and Bob Lindsay, Luis Rocha, Ted Belding, Leeann Fu, and Tom Bersano-Begey for their comments and suggestions.

### References

- Booker, L. B. (1982). *Intelligent behavior as an adaptation to the task environment*. Doctoral dissertation, University of Michigan, Ann Arbor, Michigan.
- Booker, L. B. (1985). Improving the performance of genetic algorithms in classifier systems. *Proceedings of International Conference on Genetic Algorithms and Their Applications*, pp. 80–92.
- Booker, L. B. (1987). Improving search in genetic algorithms. In L. Davis (Ed.), *Genetic algorithms and simulated annealing*, pp. 61–73. London: Pitman.
- Cantú-Paz, E. (1997). A survey of parallel genetic algorithms. IlliGAL Report No. 97003, University of Illinois, Urbana, Illinois.
- Craighurst, R. and Martin, W. (1995). Enhancing GA performance through crossover prohibitions based on ancestry. *Proceedings of the 6th International Conference on Genetic Algorithms*, Morgan Kaufmann, pp. 130–135.
- Das, R. and Whitley, L. D. (1991). The only challenging problems are deceptive: Global search by solving order 1 hyperplanes. *Proceedings of the Fourth International Conference on Genetic Algorithms*, Morgan Kaufmann, pp. 166–173.
- De Jong, K., Spears, W., and Gordon, D. (1994). Using Markov chains to analyze GAFOs. *Foundations of Genetic Algorithms 3*, pp. 115–137.
- Deb, K. and Goldberg, D. E. (1989). An investigation of niche and species formation in genetic function optimization. , *Proceedings of the 3rd International Conference on Genetic Algorithms*, Morgan Kaufmann, pp. 42–50.
- Dobzhansky, T. (1937). *Genetics and the Origin of Species*. (Reprinted 1982). New York: Columbia University Press.
- Eshelman, L. J. and Schaffer, J. D. (1991). Preventing premature convergence in genetic algorithms by preventing incest. *Proceedings of the 4th International Conference on Genetic Algorithms*, Morgan Kaufmann, pp. 115–122.
- Fernandes, C., Tavares, R., Munteanu, C. and Rosa, A. (2001). Assortative mating in genetic algorithms for vector quantization problems. *Proceedings of the 2001 ACM Symposium of Applied Computing*, pp. 361–365.
- Forrest, S. and Mitchell, M. (1993). Relative building block fitness and the building block hypothesis. *Foundations of Genetic Algorithms 2*, pp. 109–126.
- Freeman, S. and Herron, J. C. (1998). *Evolutionary Analysis*, Prentice Hall, 1998.

- Goldberg, D. E. (1989). *Genetic Algorithms in search, Optimization, and Machine Learning*. Reading, MA: Addison Wesley.
- Hillis, W. D. (1992). Coevolving parasites improve simulated evolution as an optimization procedure. *Artificial Life II*, volume X of Santa Fe Institute Studies in the Sciences of Complexity, Addison-Wesley, pp. 313-324.
- Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press.
- Holland, J. H. (2000). Building blocks, cohort genetic algorithms, and hyperplane-defined functions. *Evolutionary Computation*, 8(4): pp. 373-391.
- Huang, C-F. (2002a). *A Study of Mate Selection in Genetic Algorithms*. Doctoral dissertation. Ann Arbor, MI: University of Michigan, Electrical Engineering and Computer Science.
- Huang, C.-F. (2002b). A Study of Mate Selection Schemes in Genetic Algorithms—Part II. *Evolutionary Computation*, in preparation.
- Koza, J. (1992). *Genetic Programming*. MIT Press.
- Mahfoud, S. W. (1995). *Niching Methods For Genetic Algorithms*. Ph. D. thesis, IlliGAL Report No. 95001, University of Illinois, Urbana, Illinois.
- Mayr, E. (1942). *Systematics and the Origin of Species*. (Reprinted 1982). New York: Columbia University Press.
- Miller, R. G. (1986). *Beyond ANOVA, basics of applied statistics*. John Wiley & Sons.
- Mitchell, M. (1996). *An Introduction to Genetic Algorithms*. Cambridge, MA: MIT Press.
- Nix, A. E. and Vose, M. D. (1992). Modelling genetic algorithms with Markov chains. *Annals of Mathematics and Artificial Intelligence* #5, pp. 79-88.
- Potter, M. A., De Jong, K. A., and Grefenstette, J. J. (1995). A coevolutionary approach to learning sequential decision rules. *Proceedings of the 6th International Conference on Genetic Algorithms*, pp. 366-372.
- Price, P. W. (1996). *Biological Evolution*. Saunders College Publishing, Harcourt Brace College Publishers, New York, NY.
- Rosin, C. D. and Belew, R. K. (1997). New methods for competitive coevolution. *Evolutionary Computation*, 5(1): pp. 1-29.
- Roughgarden, J. (1979). *Theory of Population Genetics and Evolutionary Ecology*, Prentice-Hall, 1979.
- Rudolph, G. (1994). Massively parallel simulated annealing and its relation to evolutionary algorithms. *Evolutionary Computation*, 1(4): pp. 361-383.
- Russel, P. J. (1998). *Genetics*. Benjamin/Cummings, 1998.
- Spears, M. W. (1994). Simple subpopulation schemes, *Proceedings of the 1994 Evolutionary Programming Conference*, World Scientific. pp. 296-307.
- Suzuki, J. (1993). A Markov chain analysis on a genetic algorithm. *Proceedings of the 5th International Conference on Genetic Algorithms*, pp. 146-153.
- Todd, P. and Miller, G. (1991). On the sympatric origin of species: mercurial mating in the quicksilver model. *Proceedings of the 4th International Conference on Genetic Algorithms*, Morgan Kaufmann, pp. 547-554.
- Todd, P. and Miller, G. (1997). Biodiversity through sexual selection. *Artificial Life V: Proceedings of the Fifth International Workshop on the Synthesis and Simulation of Living Systems*, Cambridge, MA: MIT Press/Bradford Books, pp. 289-299.
- Vose, M. D. and Liepins, G. E. (1991). Punctuated equilibria in genetic search. *Complex Systems* 5, pp. 31-44.